

UNITED STATES PATENT APPLICATION FOR

CLIENT AWARE CONTENT SCRAPING AND AGGREGATION
IN A WIRELESS PORTAL SYSTEM

Inventors:

SATHYA KAVACHERI

THOMAS MUELLER

LUU TRAN

Prepared by:

WAGNER, MURABITO & HAO, LLP

Two North Market Street

Third Floor

San Jose, California 95113

(408) 938-9060

CLIENT AWARE CONTENT SCRAPPING AND AGGREGATION IN A WIRELESS PORTAL SYSTEM

CROSS REFERENCE TO RELATED APPLICATION

5 This patent application is related to co-pending patent application serial number _____, filed on _____, by Luu Tran et al., entitled " Client Aware Detection in a Wireless Portal System", attorney docket No.: SUN-P6087, which is hereby incorporated herein by reference in its entirety.

10 FIELD OF THE INVENTION

 The present claimed invention relates generally to the field of wireless communication systems. More particularly, the present claimed invention relates to client aware content scrapping and aggregation in a client independent wireless environment.

15 BACKGROUND ART

 The Internet has become the dominant vehicle for data communications. And with the growth of Internet usage has come a corresponding growth in the usage of Internet devices, wireless devices and services.

20 The growing base of Internet users has become accustomed to readily accessing Internet-based services such e-mail, calendar or content at any time from any location. These services, however, have traditionally been accessible primarily through stationary PCs. However, demand is now building for easy
25 access to these and other communication services for mobile devices.

As the demand for mobile and wireless devices increases, enterprises must rollout new communication capabilities beyond the reach of traditional wired devices, by extending the enterprise with extra-net applications, etc., to effectively and efficiently connect mobile employees with their home base. As the number of digital subscribers grows, traditional wireless providers must find applications suitable to the needs of these new mobile users.

However, service providers are not the only ones seeking applications to meet the growing service needs of wireless users. Traditional portal developers are also extending their traditional PC browser desk-top services to these new wireless markets.

With the growth of the wireless market comes a corresponding growth in wireless business opportunities, which in today's ever-growing markets means, there is a plethora of services available to customers of the enterprises that use these services. Many wireless service providers are now looking to add to basic core services by extending services such as e-mail, short messaging service notification, and other links to Internet Protocol (IP) based applications to drive additional business and revenues.

As the wireless market grows and Internet access becomes more mainstream and begins to move to new devices, wireless service providers are looking to develop highly leveraged Internet Protocol based applications on top of existing network infrastructure. To meet the growing demand for wireless client devices, enterprises need to provide access to any type of service from any type of device from anywhere and to provide content suitable for these devices without incurring substantial cost overhead.

The growth in wireless devices also means that traditional computer users who used to be tied to their desktop computers may now be mobile and would require remote access to network applications and services such as email. The mobility of wireless users presents a host of challenges to service providers who may have to provide traditional service to these new wireless devices. One such service is provided by Sun Microsystems, Inc., through its iPlanet™ platform to allow service providers to grow their services from basic traditional services such as voice to leading edge wireless applications with carrier-grade reliability and performance.

In addition to the traditional network applications that these new wireless users seek the growth of the Internet and the introduction of new Internet enabled wireless devices have led to the explosive use of community-based web site or portals. This growth in portals has created a need for wireless environments to provide portal support to handle the collection of data related to different topics such as news, stock quotes, applications and services required by wireless device users.

Figure 1 depicts a prior art wireless client dependent based environment solution to handle similarly configured wireless client running similar applications or portals. The environment depicted in Figure 1 includes wireless devices such as a Wireless Applications Protocol (WAP) phone 101, a wireless PC 102, a refrigerator 103, etc. In general, the wireless environment depicted in Figure 1 is categorized into the network (Internet 104), Clients (e.g. mobile phone 101, PCs 102 and household appliances 103) and resources (e.g., web-sites 105, portals 106 and other applications 107).

For most of the wireless clients connected to the Internet 104, portals 106 offer the client the starting point of experiencing the Internet 104. Portals 106 are typically community-based web-sites that securely hold a collection of data
5 related to different topics, including such applications as news, stock quotes, etc. For example, a wireless client connecting to the Internet will first login to a web portal site (e.g., yahoo) and from there browse through various sites to search for a host of different services.

10 The portals typically reside in a portal server which bundles an aggregation of services provided by an Internet service provider and provide these services to wireless clients. A wireless portal server such as that developed by Sun Microsystems, Inc. provides such portal access to wireless application
resources residing on resource servers A 108, B 109 and C 110.

15 The prior art wireless server depicted in Figure 1 primarily supports the two major types of browsers known by most Internet users. These include the Microsoft Internet Browser and the Netscape Communicator Browser. These browsers are both HyperText Markup Language (HTML) based and suitable for
20 some wireless devices, especially devices with large display screens. However, as wireless display screens get smaller in size, traditional HTML browsers are no longer suitable for transmitting content to these wireless devices.

25 To ensure suitable content delivery, wireless device and wireless software providers have developed a myriad of micro-browsers that appropriately adapt to these wireless devices with different display screen requirements in order to take advantage of the numerous contents on the Internet. The availability of

these new micro-browsers means that service providers do not have to create different sets of content for different wireless devices even if the devices are dissimilar.

5 In the prior art system depicted in Figure 1, content delivery is primary performed in HTML over such protocols as TCP and HTTP. This content is suitable for devices with wider screen display such as desktop or wireless PCs. Due to the large screen displays these prior art systems have, a resource server does not have to format content prior to delivery to these devices.

10 Thus, for the prior art server to support the various types of wireless client that service provider's support, a system's administrator has to painstakingly manually add client specific content to support each wireless client if and when the client attempts to access the server.

15 Content delivery by a service provider in the prior art system depicted in Figure 1 is typically conducted by syndication. In this process, the service provider is able to gather content from multiple web-sites and aggregate (syndicate) the content to give the appearance that the service provider is the original source of the syndicated content. In this prior art content syndication method, the syndicated content is only suitable for wireless devices of the same or similar configuration.

20 Content scrapping from multiple sites look fine in most prior art cases except for devices that require strictly formatted content, e.g., wireless phones or hand-held devices. For HTML devices, the service provider can strip off some well known markup tags and modify links embedded in the Uniform Resource

locator (URL) to reference URL resources through the service provider's system. However, as wireless devices proliferate and the configurations of these devices differ, the prior art method of syndicating content scrapped from multiple web-sites, while giving the appearance that the content is emanating from a single
5 site, for presentation to these devices is unable to provide suitable formatted content for such devices. This is because many of the sites being scrapped do not have the ability or capability to provide content in a client aware manner, i.e., based on specific client type information.

10 Another drawback of the prior art scrapping system is that content scrapping is typically performed only for desktop computer systems. In these systems, the URL scrapper retrieves pages specified by the URL and inserts the pages into a table cell defined for the content scrapped. The retrieved pages are typically HTML pages which contain entire pages including corresponding
15 <HTML> <HTML> tags. In these instances, the desktop browser either ignores the tag syntax or is able to understand the syntax. However, when a page is retrieved based on any wireless markup language, duplication of the wireless tags causes a syntax error and the page is unable to be displayed.

20 As the number of models of wireless clients increase, having restricted content and very limited information about client characteristics impairs the ability of service providers to take advantage of new wireless technologies and provide efficient and cost effective services. This also impairs the ability of the wireless client to enjoy the full richness and look and feel of user interfaces
25 provided by the server.

SUMMARY OF INVENTION

Accordingly, to take advantage of the myriad of wireless applications and the numerous wireless clients being developed, a wireless server with extensibility capabilities to allow content delivery to wireless clients to be
5 configured based on pre-defined client type and formatted by the wireless server is needed. A need exists for "out-of-the-box" wireless client aware system solutions to allow technically unsophisticated end-users to connect to the wireless environment without unduly tasking the end-user's technical abilities. A need further exists for an improved and less costly device independent
10 system, which improves efficiency and provides content to various wireless clients of different configurations without losing the embedded features designed for these devices.

A wireless portal system has a wireless server with a client aware content
15 scrapping and aggregation method and system. The client aware content scrapping and aggregation system includes logic for identifying client wireless devices connecting to the wireless by using particular characteristics of the client in presenting content selected in response to connection requests from the client to the server. In one embodiment of the invention, the client aware content
20 scrapping system receives an indication of a client type and uses this information to automatically search for content from a variety of web-sites on the Internet and formats content selected by the end-user for delivery to the client. The scrapping logic further allows content gathered from the various web-sites selected by the end-user to be aggregated in a format suitable for delivery to the
25 client. A link re-writing logic rewrites embedded in the content gathered during the scrapping procedure of the present invention in order to make the aggregated content appear stored in the wireless server. Content selection in

the present invention is user defined, but content delivery is defined by the scrapping logic, which allows the user selected content to be delivered in a client aware manner based on the client type information.

5 Embodiments of the present invention are directed to a system and a method for wireless client aware content selection and presentation in a wireless network environment. In general, embodiments of the present invention vary the degree of providing content gathered from a variety of web-sites to a wireless client connecting to the wireless environment. The present invention
10 implements client aware content aggregation to retrieve content based on detailed client type information. In other words, the invention provides client specific content aggregation of content scrapped from various web-sites and presented in a wireless network environment. The invention is suitably adapted to function in a wireless portal environment.

15 Embodiments of the invention include an integrated Internet content management solution designed to improve the availability of Internet content gathered from a variety of web-sites for delivery to a variety of differently configured wireless clients based on client type information specified by the
20 client to a wireless server. The present invention allows for the intelligent aggregation of Internet content gathered from dissimilar web-sites or geographically dispersed content into a coherently and cohesively formatted content suitable for delivery to the requesting client.

25 Embodiments of the invention include a content scrapper to dynamically search back-end services in the wireless environment for end-user specified content for delivery to the client's home page in the wireless server. The content

scrapper further creates a portal session based on the client type information provided by a client connecting to the wireless network. Information created in the session is used to retrieve client specific content from back-end services connected to the wireless server.

5

In one embodiment of the present invention, the content scrapper includes provider logic which generates Wireless Markup Language (WML) decks and other wireless adaptable language decks based on the method invoked to provide content to a requesting client. The provider logic generates the correct WML that is appropriate for the type of Wireless Application Protocol device requesting information based on pre-existing knowledge of the supported device type.

10

In another embodiment of the present invention, the content scrapper logic is responsible for encapsulating or formatting the back-end server content into appropriate markup language for delivery to the wireless clients. Service applications in the wireless server of the present invention know which scrapping method to invoke based on information embedded in the User Agent contained in the HTTP header from a client request.

15

20

Embodiments of the present invention further include client aware content source logic that creates a list of the Uniform Resource Locators used by the scrapping logic to gather content specified by the end-user. In other words, the client aware source logic acts as an index to the URLs that points to the websites where content is scrapped. Content scrapping is based on the type of client device, the type of services offered by the service provider and whether the client is authorized to scrap the content. An Application Programmable

25

Interface can be used to collect extensible data sets that include custom client parameters for automatically identifying content of a particular client or class of clients.

5 Embodiments of the present invention further include content aggregation logic that is capable of dynamically aggregating content scrapped from multiple sources into cohesive content for delivery to clients connecting to the wireless server. In the present invention, the service provider can remotely control the content aggregation logic in order to dynamically make content
10 available to the client depending on the client type information stored in the wireless server.

 In one embodiment, automatic client detection logic identifies the type or class of the client and stores this information into a client session data structure.
15 The client session information can then be used by the content selection system to automatically access the most pertinent content data for the client using an intelligent file retrieval system. Client identification or class information can be used in automatically determining whether content desired by the client is available, selectable and presentable to the client. The content selection system
20 receives an indication of a client type and uses this information to automatically construct content available for the client. The wireless server also modifies the links that are in the scrapped content so that the client must go through the wireless server to get to the link.

25 These and other objects and advantages of the present invention will no doubt become obvious to those of ordinary skill in the art after having read the

following detailed description of the preferred embodiments which are illustrated in the various drawing figures.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of this specification, illustrates embodiments of the invention and, together with the description, serve to explain the principles of the invention:

5

Prior Art Figure 1 is a block diagram of a conventional device dependent wireless system;

10 Figure 2 is a block diagram of an implementation of a device independent wireless system of an embodiment of the present invention;

Figure 3 is a block diagram of an exemplary internal architecture of the wireless server of Figure 2;

15 Figure 4 is a block diagram of an embodiment of the wireless data scrapping environment of the present invention;

20 Figure 5 is a block diagram of an embodiment of an internal architecture of a wireless Uniform Resource Locator (URL) of an embodiment of the present invention; and

Figure 6 is a flow diagram of an exemplary implementation of the scrapping procedure of one embodiment of the present invention.

25

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Reference will now be made in detail to the preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings.

While the invention will be described in conjunction with the preferred

5 embodiments, it will be understood that they are not intended to limit the invention to these embodiments.

On the contrary, the invention is intended to cover alternatives, modifications and equivalents, which may be included within the spirit and
10 scope of the invention as defined by the appended Claims. Furthermore, in the following detailed description of the present invention, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be obvious to one of ordinary skill in the art that the present invention may be practiced without these specific details. In other
15 instances, well-known methods, procedures, components, and circuits have not been described in detail as not to unnecessarily obscure aspects of the present invention.

The invention is directed to a system, an architecture, subsystem and
20 method to manage a wireless clients content selection in a client independent wireless environment in a way superior to the prior art. In accordance with an aspect of the invention, a wireless server provides wireless client aware content selection and presentation which enables client characteristics of devices to be used in determining the type of content to present by the wireless server.

25

In the following detailed description of the present invention, a system and method for a wireless Internet protocol based communication system are

described. Numerous specific details are not set forth in order to provide a thorough understanding of the present invention. However, it will be recognized by one skilled in the art that the present invention may be practiced without these specific details or with equivalents thereof.

5

Generally, an aspect of the invention encompasses providing an integrated wireless Internet server which provides a wide range of voice, data, video and other services to wireless clients which may connect to the wireless environment to be serviced alongside predefined wireless clients. The invention
10 can be more fully described with reference to Figures 2 through 6.

Figure 2 depicts a wireless device independent based environment of the present invention. The wireless environment depicted in Figure 2 comprises a wireless application protocol (WAP) based phone 201, a WAP transmission
15 infrastructure 202, a WAP gateway 205, the Internet 206 and a wireless server 210. The WAP gateway 205 typically resides on the Local area network (LAN) within a telecom carrier premise. It is generally not a part of the wireless server 210. The WAP gateway 205 is responsible for converting the Wireless Markup Language (WML)/Hyper Text Transport Protocol (HTTP) content and protocol
20 into a binary compressed, encoded, encrypted version of WML over WAP.

Conversely, the WAP gateway 205 also performs the translation of WAP commands into HTTP requests that can be sent over the public Internet 206. For example, in a GSM network, when a phone transmission is received by the
25 mobile switching center, the gateway 205 distinguishes the transmission as a packet data and sends it to the proper channel to be processed. The WAP gateway 205 decompresses and decrypts the packets, as well as several other

functions and formats the data into an HTTP request that is sent to the wireless server 210. The WAP gateway 205 can also store user's bookmarks, two of which could point to the wireless server's messaging and other resource services. The wireless server 210 communicates Wireless Markup Language (WML) over HTTP on the front -end and communicates in native protocol of the target server on the back-end.

The wireless server communicates to these back-end resource servers using the backend server's native protocol. For example, the wireless server may communicate to resource server A 211 which may be a messaging server using IMAP. Lightweight Directory Access Protocol (LDAP) is used for all communications to and from the resource server B 212. And an Extensible Markup Language (XML) protocol may be used to communicate with resource server C 213.

Although the wireless server depicted in Figure 2 is capable of communicating in these native protocols shown in Figure 2, the wireless server's 210 protocol handling capability can be extended to support a variety of other protocols. The wireless server implements the WML interface and generates the corresponding WML content based on what it receives from the back-end server. The wireless server 210 also processes incoming HTTP requests in which a wireless device sending data or a request to the back-end servers. The wireless environment depicted in Figure 2 typically supports wireless devices of dissimilar configuration and is thus device independent.

Figure 3 is a block diagram illustration of one embodiment of the wireless server of the present invention. Wireless Server 210 (WS) comprises, Content

Provider Service module 310, Client Detection module 315, Uniform Resource Locator (URL) 320, Session Service (SS) module 330, Client Data module 340 and Profile Service (PS) module 350. The server 210 also couples to back-end service 360 to retrieve content requested by the wireless clients connecting to the server 210.

The wireless server depicted in Figure 3 is a flexible, scalable, extensible and capable of supporting a rich evolving range of networks such as Global System for Mobile communication (GSM) Networks, Code Division Multiple Access (CDMA) Networks, Time Division Multiple Access (TDMA) Networks, Third Generation (3G) Networks and others.

The architecture of the server is also capable of handling a variety of wireless environments and markup languages such as the wireless markup language (WML), the handheld device markup language (HDML) and the hypertext markup language (HTML). The server is capable of providing support for multiple devices and is easily adaptable and extensible to additional devices and markup languages.

Still referring to Figure 3, Provider Service module 310 is coupled to provide a client with the ability to select and present content to a user's homepage and navigation links to other providers in the server 210. An example of how an end-user can select content using the Provider Service 310 is described in co-pending U.S. patent application titled " Client Aware Content Selection and Retrieval in a Wireless Portal System, file on even date, Serial No.: _____, assigned to the assignee of the present invention and hereby incorporated by reference herein.

URL scrapper 320 is coupled to the Provider Service 310 and the Client Detection 315 to store URLs for all the defined client types of clients supported by the wireless server 210 in Profile Service 340. URL scrapper 320 parses incoming client requests to determine and extract client type information in order to deliver the corresponding scrapped content to the client in a format that is customized for the client. In the present invention, the profile attributes and the URLs for each supported client type is configured by the systems administrator of the wireless server 210 and not by the end user. Client type information may be dynamically added during system run-time to allow newly authorized client types to connect to the server 210. URL Scrapper 320 enables the service provider to syndicate multiple content scrapped from a variety of web-sites into a single source for presentation to the wireless clients connecting to server 210. In the present invention, content is scrapped from the variety of web-sites in a client aware manner. In other words, content is scrapped in response to pertinent unique client information provided by the client.

In the scrapping procedure of the present invention, wireless server 210 acts as an intermediary server for URL Scrapper 320 requests to the different web-sites from which content is gathered. Client requests to URL 320 are handled by the wireless server 210 which communicates with the back-end resource service 360 to gather content from the various resources stored in the back-end resource service 360. Requests to the back-end service 360 are processed based on the client type information provided by the Client Detection service 315 of the wireless server 210. The Function of Client Detection Module 315 is described in the co-pending US Patent Application entitled "CLIENT AWARE DETECTION IN A WIRELESS PORTAL SYSTEM", filed on even date,

Serial No.: _____, assigned to the assignee of the present invention and hereby incorporated herein by reference.

In the scrapping environment of the present invention, when the URL
5 Scrapper 320 has to provide content to multiple clients with different configurations, for example as depicted in Figure 4, a wireless phone 201c, a hand-held device 201b, a wireless PC 201d or a house-hold appliance 201a, etc., the wireless server 210 acts as a switch by routing requests from the various clients to the appropriate back-end resource service, e.g., web-sites 360a - 360c
10 depending on the client type. Therefore, the wireless server 210, using the URL Scrapper 320, could act as a proxy server if the scrapped web-site does not have client awareness capabilities.

URL Scrapper 320 uses the Client Data 340 to access a file path property in
15 the server 210 to retrieve appropriate device specific templates. In the present invention, client data objects are extensible to allow additional properties to be added as needed for use by URL Scrapper 320. The present invention provides methods to display specific content in the form of channels. On an HTML device, these channels appear as table cells and on WML devices the channels
20 appear as links to WML cards containing the contents of the channel.

Reference is now made to Figure 5 that is a block diagram illustration of one embodiment of the URL Scrapper 320 of the present invention. As depicted in Figure 4, URL Scrapper 320 comprises a client aware content source 410, client
25 aware content aggregator 420 and client aware link rewriter 430.

Client aware content source 410 is coupled the content aggregator 420 to receive clients request to URL Scraper 320 and search web-sites for content specified in the User Agent in the client's request header from target back-end resource services. The content located by content source 410 is then passed to
5 content aggregator 420 for processing.

Content Aggregator 420 is coupled to content source 410 to parse and aggregate content scrapped from the various web-sites 360a - 360c, as depicted in Figure, identified by content source 410. In the present invention, Content
10 Aggregator 420 combines the various content scrapped from the numerous web-sites into a coherent and cohesive client aware content suitable for delivery by the wireless server 210 to clients 201a - 201d connecting to the server 210.

In processing the content combination, Content Aggregator 420 uses
15 Content Link Rewriter 430 to rewrite links embedded in the scrapped content to make the client aware content that is ultimately provided to the wireless client appear to originate from the wireless server 210. All links therefore go through the wireless server 210. By rewriting links, the Content Link Rewriter 430 ensures that the client does not have links embedded in the content provided by
20 the server 210. Content Link rewriter 430 also enables the service provider to customize content scrapped from other sites to make the content presentable to its clients. For example, the service provider may remove advertisements from content scrapped from a web-site when the content is presented to the service provider's clients.

25
Reference is now made to Figure 6 which is an exemplary flow diagram of one embodiment of the scrapping procedure of the present invention. The

scraping procedures commences at step 600 when the wireless client 201a -
201d initiates a content request to server 210. At step 610, the server access the
client's request to scrape information from a resource server 360a - 360c. The
wireless server 210 access the client's request based on the client type
5 information stored in the client detection service 315 at step 620.

At step 630, the URL Scraper 320 selects the appropriate back-end
resource service based on the accessed client type information. The URL
Scraper 320 then uses the client type information to scrap data from the selected
10 back-end resource service based on the content specified in the User Agent
presented in the client's request header at step 640.

At step 650, the URL Scraper 320 aggregates the scrapped data and
rewrites links contained in the scrapped data so that the links go through the
15 wireless server 210 as a proxy to the back-end resource server 108 - 110. After
rewriting the links contained in the scrapped data, the URL Scraper 320 then
provides the scrapped data in a format that is customized for the client's display
capabilities at step 660. The URL Scraper 320 then ends the scraping process
at step 670.

The foregoing descriptions of specific embodiments of the present
invention have been presented for purposes of illustration and description. They
are not intended to be exhaustive or to limit the invention to the precise forms
disclosed, and obviously many modifications and variations are possible in light
25 of the above teaching. The embodiments were chosen and described in order to
best explain the principles of the invention and its practical application, to
thereby enable others skilled in the art to best utilize the invention and various

embodiments with various modifications are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the Claims appended hereto and their equivalents.

20010618/ACM/DKA